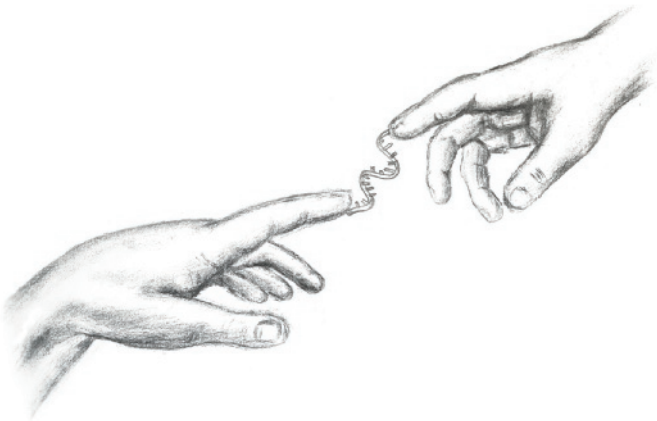
CHAPTER 1

General introduction



S.B. van der Meulen[1,2] and J. Kok[1,2]

[1] Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands
[2] Top Institute Food and Nutrition (TIFN), Wageningen, The Netherlands

Partially published in FEMS Microbiology Reviews, **41**:S220–S243

Regulation is a key feature of a successful modern and complex society in which we humans live. Without it we would face a difficult job maneuvering in current day traffic and could not rely on public facilities paid by the government through our tax system. Looking at ourselves as multicellular organisms, we are also "planned" in advance by our genetic code and the expression of our genes is highly regulated to facilitate cell differentiation. This is how we develop various types of tissues and organs that originated from only one totipotent cell after fertilization. Single cellular organisms like bacteria are usually not able to drastically change their morphology in the way higher organisms can. Nevertheless, some bacteria like the soil bacterium *Bacillus subtilis* can differentiate into spores when certain nutrients run out in their environment (1). Others produce flagella to become motile (2) and some bacteria make pilins to exchange DNA (3). All bacteria have in common the fact that their growth can be separated into different stages, they can cope with stressful environmental conditions and can utilize various molecules for metabolic purposes. To be always prepared for all conditions and states a cell could possibly encounter would require energy and chemical investments that would simply be too costly for a single bacterial cell. Therefore, bacteria (but basically any cell from any organism) have developed mechanisms to regulate gene expression. These systems ensure that genes are switched on or off whenever the cell requires, or not, the particular function that a gene codes for. Genes are not only regulated by on/off-type switches, gene expression can also be constitutive or gradually vary. In some cases, genes are heterogeneously expressed in cells in a population of which the genetic background is identical (4, 5). If the underlying mechanism for this variation is a risk spreading strategy that has evolved in an unpredictable changing environment, one can speak of bet-hedging (6). Sporulation is an example of bet-hedging; it enables some cells to sporulate while others continue to grow in the current environment (7). The response of bacteria to growth conditions, nutrient availability or to sometimes rapidly changing environmental stimuli can occur at the DNA, RNA and/or protein level. Relatively recently, tremendous insights have been gained into these regulatory processes by deploying various types of '-omics' techniques. The nucleotide sequencing of entire genomes, or genomics, marked a revolution in molecular biology. After obtaining an overall picture of all the genes in genomes, the next step was to study the expression of all these genes at a given time point, or transcriptomics. These studies have been boosted by advances in DNA micro-array technology and RNA deep-sequencing. The latter has now resulted in the discovery in bacteria of many new types of RNA species, including small regulatory RNAs and RNAs that target foreign nucleic acids derived from bacteriophages and plasmids. The high-throughput sequencing results in an enormous amount of data, requiring proper storage and intelligent algorithms to process this data in order to find the embedded biological information.

Regulatory mechanisms in bacteria

The central dogma in biology describes how genes encoded on DNA are transcribed into

messenger RNA, which is subsequently translated into proteins. Hierarchically, overall bacterial gene expression by transcription is mostly controlled by sigma factors that, together with the RNA polymerase holoenzyme, form the RNA polymerase complex. Sigma factors recognize specific DNA sequences in the region upstream of promoter and transcription start site. Seven sigma factors have been identified in *Escherichia coli*. The protein σ^{70} (RpoD, or housekeeping sigma factor) is responsible for transcribing most of the genes, while for example RpoS is involved in the general stress response (8). Other sigma factors have been linked to specific stress conditions such as extracytoplasmic stress (RpoE) (9) or heat (RpoH) (10), while the regulons of some sigma factors functionally overlap (11). Another major group of protein regulators of gene expression are transcription factors (TFs). A TF is a DNA binding protein that recognizes a specific motif in the promoter region and by binding to it controls the downstream gene(s) through activation or repression of transcription. At the level of DNA, methylation of an adenine or cytosine base can also influence transcription. Usually, the presence of methylated nucleotides in promoter sequences results in a decrease of transcription (12). Fairly recently, the widespread occurrence and importance in bacteria of a whole set of RNA molecules with regulatory functions was recognized (13). **Table 1** gives an overview of several bacterial RNA species of which those that are involved in gene regulation will be discussed in more detail below.

Small regulatory RNAs. Particularly in the last decade, hundreds of small regulatory RNAs (sRNAs) have been identified in many bacterial species. sRNAs now greatly outnumber the protein regulators specified by bacterial genomes, not only in absolute number, but also in diversity of function (14). sRNAs are small transcripts that are very heterogeneous in size, ranging from around 50 to up to ~350 nucleotides. sRNA genes are typically located in intergenic regions (see **Figure 1A**), are controlled by orphan promoters and have Rho-independent terminators. These features have previously been used to predict sRNA genes in bacterial genomes (15). sRNAs act post-transcriptionally in the regulation of messenger RNAs (mRNAs) by means of base pairing. In 1983, Mizuno and co-workers reported the first chromosomally encoded sRNA, MicF (16). MicF regulates the mRNAs *ompF* and *ompC* that encode essential outer membrane proteins in *E. coli* by blocking the ribosome binding site and thereby translation. A follow-up publication by the same researchers suggested that regulation by RNAs might be a general form of regulation in bacteria and eukaryotes (17). This assumption turned out to be very correct considering the subsequent discovery of a host of regulatory RNAs in all organisms in which their presence has been examined (18).
Bacterial sRNAs mostly target mRNAs; they can modulate gene expression via four different molecular mechanisms (see **Figure 1B**). RNA stability can be influenced such that the target mRNA is repressed by RNase cleavage (19, 20), whereas activation is mediated through protection by the sRNA of an RNase cleavage site (21). By binding of the sRNA and thereby blocking the ribosomal binding site (RBS) or Shine Delgarno (SD) sequence, translation of

the mRNA cannot start (17). Translation can be activated by an sRNA in the situation where the mRNA normally folds in such a way that it blocks its own RBS; the sRNA unfolds this structure by upstream base pairing.

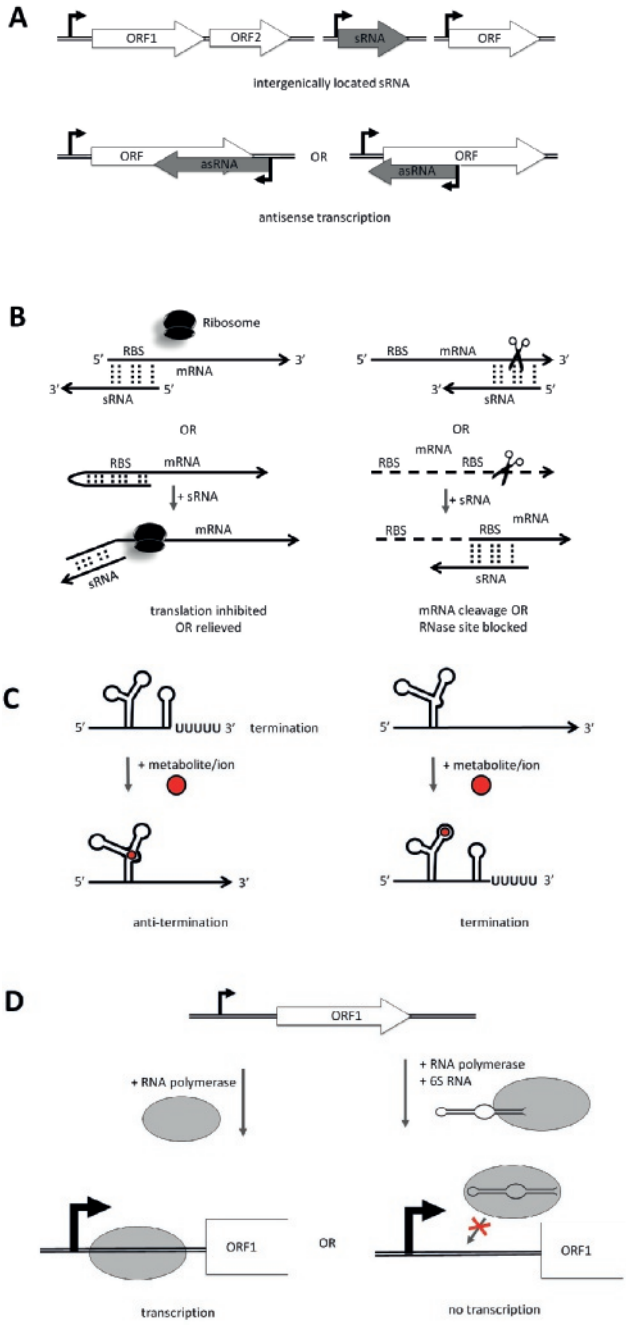


Figure 1. RNA regulators in bacteria. (**A**)**,** sRNAs are transcribed intergenically (top), antisense RNAs are transcribed from the opposite strand of a coding gene (bottom). (**B**)**,** sRNAs can influence RNA translation and RNA stability, both either with a negative or a positive effect on gene expression (see text). (**C**)**,** Riboswitches are RNA structures located in the 5′-UTR of an mRNA. The RNA structure changes upon binding of a metabolite or ion, which results in either a change in anti-termination or termination of transcription. With a similar mechanism, translation can also be affected by riboswitches (see text). (**D**)**,** The 6S RNA can form a complex with RNA polymerase, thereby preventing transcription.

Current research on sRNAs focuses on the functional characterization and mode of action of these molecules. It is now evident that base-paring between an sRNA and its target mRNA(s) plays a crucial role in the regulation process. Two RNA bases can interact optimally via G:C base pairs as these involve three hydrogen bonds. A:U and non-productive G:U pairing results in two hydrogen bonds. The strength of base pairing interaction, thus, depends on the complementarity between the two RNA species and can be predicted with sRNA target prediction software. These tools, which are often accessible online and include such examples as IntaRNA (22), TargetRNA2 (23) and RNApredator (24), calculate the minimal free energy of the sRNA-mRNA base pairing. The secondary structure of sRNAs and the potential energy that is necessary to melt pairing nucleotides is included in these calculations. A large number of false positive hits are often obtained, however, and pre-determined settings in the programs for the interaction window of the RNAs do not allow taking into account the fact that quite some variability exists in the lengths of mRNA leaders. The software CopraRNA, which uses homologous sRNA sequences from at least three species as an input, is reported to improve target predictions (25). sRNAs can distinguish their mRNA targets by an astounding precision, as is shown by the sRNA SgrS, a molecule that can discriminate a target mRNA based on a single hydrogen bond (26). This surgical precision is a reflection of the sophistication in the fine-tuning of bacterial gene regulation, and presents a serious challenge to in *silico* target predictions.

sRNAs in action. sRNAs can regulate a host of different biological processes. For instance, sRNAs have been found to play a role during stress conditions (27-29), in virulence (30, 31) and metabolism (32). sRNAs can regulate multiple mRNAs, while one mRNA can be regulated by various sRNAs (33). In some cases, sRNAs perform a crucial role in overcoming a potentially lethal stress, while in other cases fine-tuning by sRNAs enables homeostasis. The sRNA RyhB from *E. coli* provides an example of the latter in regulating iron homeostasis. Under iron limiting conditions, the ferric uptake repressor protein Fur is deactivated, which enables RyhB expression. RyhB restores iron levels by downregulating nonessential proteins that contain iron or are involved in iron storage (34). Another example of an sRNA that is controlled by a transcription factor is the bifunctional SgrS from *E. coli* and *Salmonella* species. The regulator SgrR activates SgrS expression after the uptake of toxic phospho-sugars. SgrS upregulates the phosphatase YigL, which dephosphorylates the phospho-sugars. This enables the efflux of dephosphorylated sugars and reduces the stress quickly (21). To prevent further uptake of phospho-sugars by PtsG and ManX, SgrS base pairs with the mRNAs encoding these proteins (35). In addition to its non-coding regulatory functions, SgrS also encodes the peptide SgrT, which interacts with PtsG to prevent further glucose uptake by this transporter (36).

Some sRNAs themselves are under control of sigma factors. RybB of *E. coli* and *Salmonella* is controlled by the extracytoplasmic stress sigma factor σE. Cell envelope stress such as occurs from the accumulation of misfolded outer membrane proteins can activate RybB via

σE. RybB downregulates many different mRNAs encoding outer membrane proteins in order to restore envelope homeostasis (37).

An antisense RNA is a transcript complementary to another (m)RNA molecule. This *cis*-encoded element has a region that is fully complementary to part of its target (m)RNA and, thus, both RNAs can perfectly base pair. In addition to functions described above for sRNAs that is, affecting RNA stability and translation, asRNAs can also influence transcription termination and transcription interference (38). Apart from the obvious targets, it cannot be excluded that asRNAs also target other mRNAs in *trans*. The first asRNAs identified were shown to control plasmid replication and copy numbers (39, 40), later followed by asRNAs on phage genomes and transposons (41). Currently, only a relatively small number of chromosomally encoded asRNAs have been characterized in bacteria (42). In contrast to sRNAs, only a few asRNAs have been linked to specific stress conditions. One example in *B. subtilis* was recently published, in which ethanol stress was shown to induce the expression of an asRNA against *rpsD*, encoding the ribosomal protein S4, resulting in downregulation of the small ribosomal subunit (43). The functions of many of the asRNAs discovered thus far remain enigmatic.

Recent bacterial RNA sequencing studies show that antisense transcription is very pervasive. Although the length of some asRNAs is comparable to that of sRNAs, some antisense transcripts can cover whole operons. Similarly, mRNAs can also overlap in an antisense fashion with their 3'-UTRs or 5'-UTRs. The biological function of this extensive antisense transcription is not known, but it might coordinate the expression levels of the sense transcripts and/or remove transcriptional noise (44). The detection of antisense transcripts is highly variable when comparing bacterial species (38), and even varies greatly within one species (45), suggesting that many asRNAs and their functions are yet to be discovered.

Protein-binding sRNAs. Besides regulating the expression of mRNAs, a few sRNAs have been reported to influence the function of specific proteins by mimicking the structure of other nucleic acids. The most conserved non-coding RNA currently known is 6S. Although it was already discovered in 1967 because of its high abundance (46), its function was only established in 2000 (47). The tertiary structure of 6S resembles a hairpin that interacts with the σ^{70}-holoenzyme form of RNA polymerase (see **Figure 1D**). 6S represses expression from σ^{70}-dependent promoters by mimicking an open promoter complex (48). In general, 6S is a global regulator predominantly expressed in the stationary phase and its main function is linked to survival (49).

Other well-characterized protein-binding sRNAs are CsrB and CsrC. Both bind with GGA nucleotide motifs to the RNA binding protein CsrA (50). Upon binding of CsrB or CsrC, the regulator CsrA is sequestered away from mRNAs, the translation of which would otherwise be affected by free CsrA (51).

The sRNA GlmY brings nucleic acid imitation to the next level by mimicking a homologous sRNA, GlmZ. GlmZ is processed by the RNase adapter protein RapZ, after which the former

can activate the mRNA encoding glucosamine-6-phosphate synthase. GlmY prevents GlmZ processing by binding and titrating RapZ away from GlmZ, thereby preventing GlmZ from becoming a mature and active sRNA (52).

Riboswitches. In contrast to sRNAs and asRNAs, which act *in trans* to regulate other mRNAs, riboswitches are RNA structures that act *in cis*, regulating the expression of the downstream open reading frame(s) on the same mRNA (see **Figure 1C**). Riboswitch mechanisms are based on the ability of these RNA extensions to bind specific ligands such as small metabolites or metal ions. The structural switch resulting from binding of the ligand can affect transcription termination or translation. Gene activation can be achieved by a ligand that alters the RNA structure in such a way that a terminator structure is replaced by an antiterminator. Translational activation occurs when the SD sequences and/or the AUG start codon is liberated by the ligand-induced structural change in the RNA. Gene repression is mediated by opposite mechanisms (53, 54). T-boxes are a special group of riboswitches that recognize and bind tRNAs. A T-box located in the 5'-UTR of an aminoacyl-tRNA synthetases (aaRS) measures the ratio of corresponding uncharged and charged tRNA, resulting in gene activation when this ratio is low. T-boxes are also present in some leader sequences of genes/operons encoding amino acid biosynthesis and transport (55). RNA "thermometers" are structures in RNA that can act upon a shift in temperature to modulate expression of genes located downstream on the same transcript. Both cold- and heat-responsive thermosensors have been described in some bacterial species to control expression of *e.g. dnaK* and *rpoH* (56, 57).

Ribonucleases (RNases) play a major role in overall gene expression in bacteria by actively degrading all types of RNAs. sRNAs are known to recruit RNases (58) but they are also able to block their action by base pairing with and thereby covering RNase cleavage sites in mRNAs (21). On the other hand, RNases enable the processing of sRNAs and in this way control their expression levels (59). Most RNases are specialized in cleaving RNA at specific locations. Endoribonucleases like *E. coli* RNase E or its homologue in Gram-positive bacterial species, RNase Y, cleave RNA internally (60). Double-stranded RNA formed either through intramolecular base pairing or via sRNA-mRNA interactions can be cleaved by RNase III (61). The initial endoribonucleolytic cleavage products are subsequently further degraded in a 3' to 5' direction by exoribonucleases such as polynucleotide phosphorylase (PNPase), RNase R and RNase II (59). The liberated nucleotides can be re-used by the cells. The most important players in sRNA regulation are RNase E and RNase III; they act upon base pairing between an sRNA and its target mRNA (62). PNPase, like RNase E, is able to control the amount of sRNAs. *E. coli* PNPase degrades the sRNAs RybB and MicA to a greater extent than does RNase E (63). However, in another study, absence of active PNPase was shown to destabilize sRNAs like SgrS and RyhB, suggesting that PNPase rather plays a protective role (64).

Table 1. Bacterial RNA species or RNA structures

RNA	Name and function
mRNA	Messenger RNA; contains one or more open reading frames coding for protein; can carry untranslated 5' and 3' ends
rRNA	Ribosomal RNA; protein synthesis
tRNA	Transfer RNA; amino acid carrier and recognition of cognate codon
ncRNA	Non-coding RNA; performs a role(s) in bacterial cell other than encoding proteins
sRNA	Small regulatory RNA; ncRNA that regulates other mRNAs *in trans*
asRNA	Antisense RNA, transcribed from the other strand, regulates mRNAs *in trans*
tmRNA	Transfer-messenger RNA; adds alanine-tag to unfinished polypeptides when ribosome jams, inducing proteolysis; also facilitates mRNA degradation
T-box	RNA structure capable of binding uncharged tRNAs; this interaction affects transcription or translation of downstream gene(s)
RNase P	Catalytic RNA subunit of ribonuclease P; responsible for pre-tRNA cleavage
6S	Abundant non-coding RNA; interacts with σ^{70}-RNA polymerase enzyme complex, influencing transcription
Riboswitch	Metabolite-binding RNA structure; regulates downstream gene(s) by transcription (anti)termination or by affecting translation
Ribozyme	Ribonucleic acid enzyme; RNA molecule that can act as biological catalysts
RNA thermometer	Temperature-sensitive RNA structure; affects translation of downstream gene(s) after heat- or cold-shock
crRNA and tracrRNA	Both RNAs bind the Cas protein and, together, act as an RNA-guided endonuclease cutting foreign DNA, *e.g.* from a bacteriophage or plasmid

Regulatory RNAs from animals in a nutshell

Regulation by non-coding RNAs was studied in bacteria years before their discovery in animals and plants. MicroRNAs (miRNAs), small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs) are important eukaryotic regulators that are much more similar in size and function than prokaryotic sRNAs. Animal miRNAs are ~ 22 nt in length; they are produced by two RNase III enzymes called Drosha and Dicer (65). miRNAs require the Argonaut protein to repress mRNA translation and direct mRNA decay (66). siRNAs differ in their biogenesis since they derive from Dicer cleavage activity; they share perfect complementarity to their target mRNA and, upon base pairing between siRNA and mRNA, cleavage occurs (67). Another major class of regulatory RNAs is that of the long non-coding RNAs, which are autonomously transcribed (68).

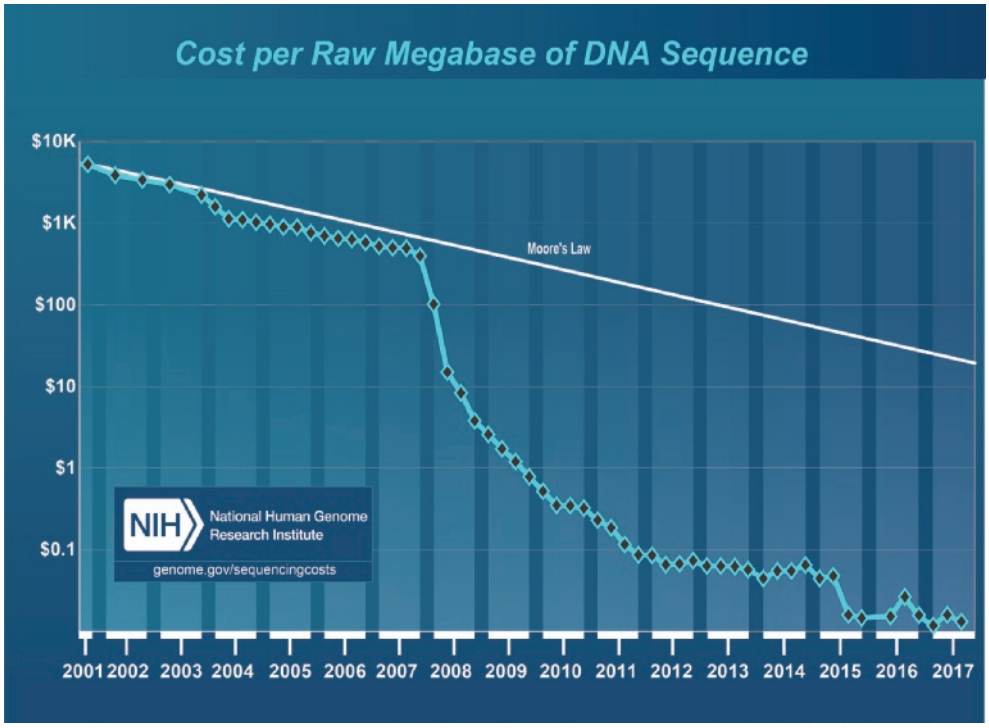


Figure 2. Overview of the development of costs per raw Megabase of DNA sequencing. The price has decreased dramatically since 2007, but has remained rather stable over the past 2 to 3 years. Source: www.genome.gov.

Sequencing and transcriptomics

A major whole-genome sequencing highlight was the first human genome unraveled by a large consortium in 2001 (69). Recent break-trough's in Next Generation Sequencing (NGS) have increased the speed, fidelity and capacity of sequencing tremendously. These developments were accompanied by a dramatic decrease in cost per sequenced nucleotide of more than 100.000-fold in the last decade (see **Figure 2**). In addition to genome sequencing, RNA sequencing, or in fact the sequencing of cDNA that is prepared from purified RNA, has boosted transcriptome research enormously.

Quantitative Polymerase Chain Reaction (qPCR) and Northern hybridization allow quantifying one or only a few transcripts at the same time. The latter methodology also provides insights in the size and degradation of individual RNAs. Based on a principle similar to that of Northern hybridization, DNA macroarrays for multiple RNA targets and, later, DNA microarray technology enabled genome-wide transcriptome analyses. The resolution of DNA microarrays was improved by using high-density tilling arrays. This technique provides a rough indication of transcript boundaries and also allows identifying unknown RNA species such as asRNAs and sRNAs. NGS entails massive parallel sequencing, provides

many advantages over DNA microarrays and, thus, it is currently the leading technology in transcriptome analysis. The benefits of RNA sequencing (RNA-seq) are that it is fast, offers single nucleotide resolution and a high dynamic range, is less noisy and does not suffer from cross hybridization issues (70).

A variety of spin-off applications have recently been developed that are all based on NGS. To enable RNA-seq, a cDNA library is prepared from the RNA to be examined. With variations in the way the cDNA library is prepared, specific enrichment of 5'- or 3'-ends of total RNA can be achieved, allowing predicting transcription start sites (TSSs) and transcript boundaries. Especially, the enrichment of primary transcripts is an extremely powerful approach to identifying new TSSs and operons. The technique takes advantage of the difference between primary and processed transcripts, such as tRNAs and rRNAs, as the latter carry a monophosphate (5'P) group at their 5'-end, while primary transcripts have a triphosphate (5'PPP) group at their 5'-ends. Treatment with the enzyme 5'P-dependent terminator exonuclease (TEX) specifically removes processed transcripts and, when combined with a non-TEX treated library, it allows differential RNA sequencing (dRNA-seq) (71, 72). Recently, a method called Cappable-seq was introduced, which also facilitates TSS profiling. As cappable-seq, in contrast to dRNA-seq, captures the primary 5'PPP transcripts directly, it enables TSS detection with even higher specificity (73). RNA can also be obtained from co-immunoprecipitation using RNA chaperones such as Hfq, allowing identifying the protein's RNA targets (74). Other techniques provide insights in the ribonucleotides that are in contact with the proteins with which they interact (75); actively translated mRNAs can be recognized by a method called ribosomal profiling (76). **Table 2** gives an overview of various methods that exploit NGS and have been applied in bacterial research. Many other exciting future applications are yet to be developed. Currently, third-generation sequencing methods are being developed; these technologies aim to sequence single DNA molecules. Since this results in longer reads, genome and transcript assembly becomes easier. Pacific Biosciences developed the Single Molecule Real Time (SMART) technology, which is very suitable for sequencing of (large) genomes. For another area of application, Oxford Nanopore Technology produced the MinION sequencer, a device no larger than a USB stick that enables mobile sequencing.

Table 2. Applications of next-generation sequencing methods

Method	*Application*	*Reference:*
DNA-seq	Genome (re)sequencing, metagenomics, cDNA sequencing for RNA-Seq	(77, 78)
- Methyl-seq	Genome-wide DNA methylation analysis to study epigenetics	(79)
- ChIP-seq	Chromatin immunoprecipitation to study protein-DNA interactions	(80)
- Tn-seq	Relative fitness of cells containing disruptive insertions by the Mariner transposon in diverse genes	(81)
RNA-seq	RNA deep sequencing for transcriptome analysis	(70)
- dRNA-seq	Differential expression of primary transcripts (obtained by TEX treatment) and total transcriptome (no TEX treatment)	(71, 72)
- tagRNA-seq	Differentially labeling of 5' RNA ends; discriminates primary from processed 5' RNA ends	(82)
- Cappable-seq	Tagging and capturing of 5' RNA ends for TSS profiling	(73)
- CLIP-seq - HITS-CLIP	Cross-linking immunoprecipitation to study protein-RNA interactions	(75)
- Ribo-seq	Sequencing of mRNA fragments that are actively translated and are, thus, ribosome-protected	(76)
- PARS	Parallel analysis of RNA structure	(83)
- FRT-seq	Flow-cell reverse transcription: amplification-free, strand-specific transcriptome sequencing	(84)
- Grad-seq	Gradient sequencing to identify RNA binding proteins and the RNAs that they bind	(85)
- Dual RNA-seq	Transcriptome analysis of both pathogen and host	(86)
- Hi-C	Chromosome conformation capture sequencing to study genome three-dimensional structure	(87)
- GRIL-seq	Global sRNA target identification by ligation and sequencing	(88)
- MAPS	MS2-affinity purification coupled with RNA-seq to identify sRNA-interacting transcripts	(89)
- DMS-seq	Dimethyl sulphate treatment of RNA followed by sequencing to evaluate RNA structures	(90)
- TTA RNA-seq	Total Transcriptome Amplification for single-cell bacterial RNA-seq	(91)
- NAD capture-seq	Identification of NAD-linked RNAs from bacteria by chemo-enzymatic capture and RNA-seq	(92)
- IP-dsRNA-seq	Immuno-precipitated double-stranded transcriptome sequencing	(93)

Lactococcus lactis

The Gram-positive, facultative anaerobic bacterium *Lactococcus lactis* is a highly characterized organism that is well appreciated for its application in the dairy industry (see **Figure 3A**). It belongs to the group of the lactic acid bacteria (LAB) and has obtained the

generally regarded as safe (GRAS) status. Dairy *L. lactis* belong to the subspecies *cremoris* or *lactis*. Currently (October, 2018), 150 *L. lactis* genomes with a variety of origins have been fully sequenced (source: NCBI). Environmental *L. lactis* strains derived from plants, animals or raw milk have more metabolic capabilities while their domesticated counterpart, isolated from a dairy environment, have been evolved more into specialists (94). Most of the strains used throughout this thesis work derive from a strain that was isolated in the 1950's from a dairy environment. This strain, *L. lactis* ssp. *cremoris* NCDO712, was plasmid- and phage-cured in 1983 (95). The resulting strain, MG1363, is currently the worldwide best characterized LAB. A derivative of MG1363, *L. lactis* NZ9000, carries in its genome the genes for the nisin sensor kinase NisK and corresponding response regulator NisR (96) and is commonly used as an overexpression host for a variety of homologous and heterologous proteins.

L. lactis is widely applied in the dairy industry for the production of a great diversity of semi-hard and soft cheeses, as well as quark and buttermilk. Its main function in milk fermentation is to convert the milk sugar lactose into lactic acid. The consequent acidification increases the shelf life of the end product and was, historically, an efficient way to store "milk" (97). In addition, byproducts of the metabolism of *L. lactis* provide texture, flavors and aromas (98, 99). From carbon and fatty acid metabolism, as well as from milk protein (casein) degradation by the organism, a whole bouquet of flavor compounds such as alcohols, aldehydes, esters, ketones and fatty acids are produced (100, 101). A mixture of such flavors distinguishes the smell and taste of a particular cheese. Apart from its use in the dairy industry, other more recent and novel applications are in the production of heterologous (membrane) proteins (102) and oral vaccines (103) and for lantibiotic modification and development of novel antibiotics (104).

The sequencing of the genome of *L. lactis* subsp. *lactis* IL1403 in 2001 (105) and of *L. lactis* subsp. *cremoris* MG1363 in 2007 (106) has greatly boosted the molecular and genetic research on this organism. *L. lactis* MG1363 contains a relatively small genome of 2.53 Mbp with a GC content of 35.8% (see **Figure 3B**) (106), in which six remnant or full prophage genomes have been identified (107). The scientific community employing *L. lactis* as a model organism has developed many genetic tools, such as a host of plasmid vectors (108), the NIsin Controlled gene Expression (NICE) system (109, 110), recombineering for deletion and mutation of short DNA sequences (111) and pSEUDO, a plasmid permitting the integration of genetic constructs in a transcriptionally silent chromosomal locus (112). For a full appreciation of the genetic amenability of *L. lactis*, please refer to a recent review on the topic by Kok *et al* (108).

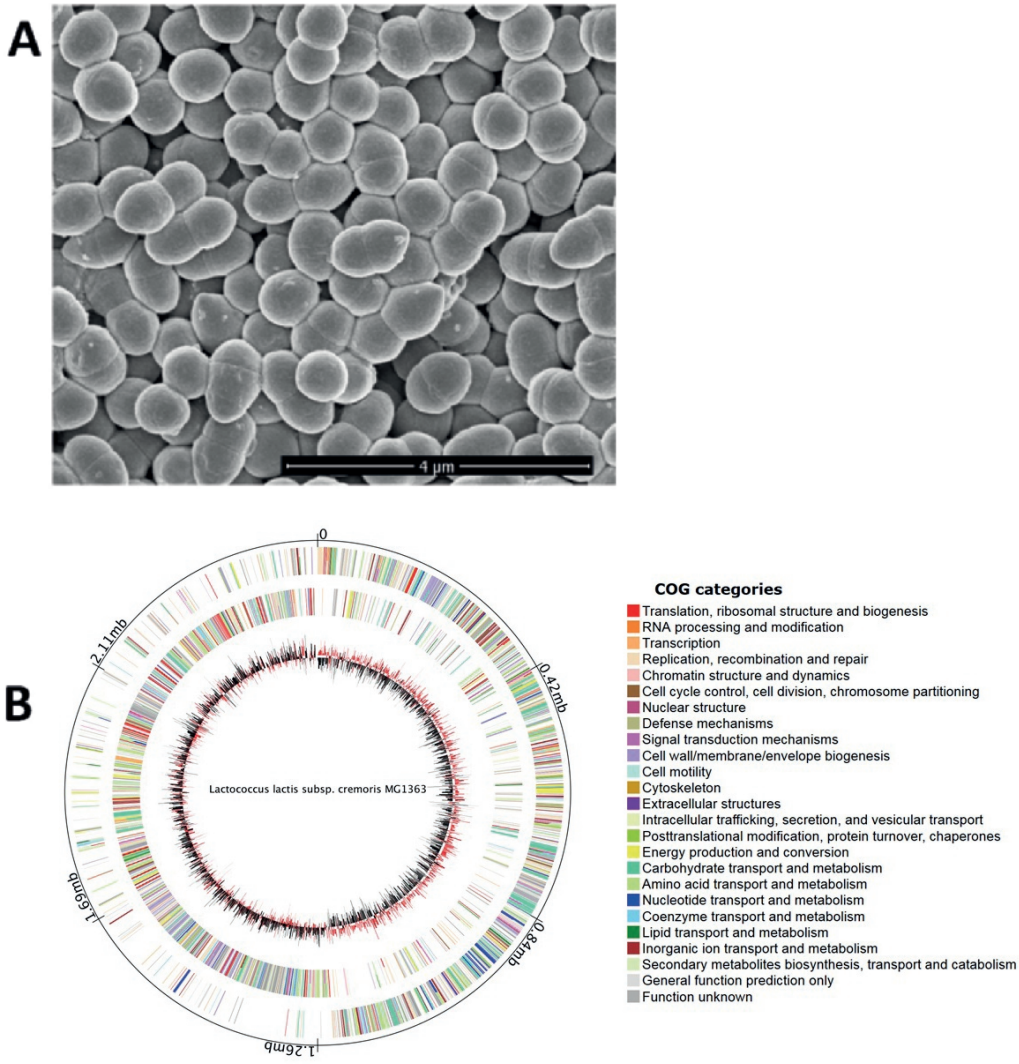


Figure 3: *Lactococcus lactis* and its genome map. (**A**)**.** Electron microscopy image of *L. lactis* subsp. *cremoris* NCDO712 cells at 25.000 x magnification. (**B**)**.** Genome map of *L. lactis* MG1363. The rings represent, from outside to inside: genome coordinates; genes on plus strand; genes on minus strand; GC skew (black); AT skew (red). The COG categories for the genes in the map were colored as shown in the legend. The map was visualized by CiVi at http://civi.cmbi.ru.nl.

Current knowledge on regulation by RNAs in L. lactis

At the start of this thesis work, nothing was known about sRNAs in *L. lactis*, let alone whether they might be involved in regulation in this organism. Since only the house keeping sigma factor (σ^{70}), and a putative sigma factor involved in extracytoplasmic function stress have been identified in *L. lactis* (113), other regulatory systems are expected to allow the

cells to adapt to the constantly changing internal and external conditions during growth. For example, the role of TFs has been very well elaborated in earlier studies in *L. lactis*. Among these protein regulators are the important global TFs CcpA and CodY, involved in carbon (114) and nitrogen metabolism (115, 116), respectively. Also, several TFs have been described with smaller regulons such as ArgR, operating in arginine metabolism (117), and CesSR, which is activated upon cell envelope stress (118). Based on sequence homology, *L. lactis* is predicted to harbor a total of 136 protein regulators (106). Many of these can be assigned to a specific pathway based on the genes encoded from the same or a nearby operon. Also, many functions of TFs can nowadays be predicted on the basis of the action of homologs in other bacteria. For the prediction of the roles that sRNAs may play in *L. lactis*, less conclusive information is available from verified homologs in other organisms, also because most of the known sRNAs are studied in Gram-negative (pathogenic) bacteria from the genus Enterobacteriaceae such as *E. coli* and *Salmonella* species. Another reason making it difficult to find homologs is the fact that limited sequence homology exists between sRNAs, due to the high variation in bacterial non-coding genome sequences. In addition, to determine the function of sRNAs, target predictions based on complementarity between the sRNA and mRNAs are not reliable, as discussed above. The only exceptions are riboswitches, which are very well conserved regulatory RNA elements albeit that some ligands can be trapped in RNA aptamers that are similar in structure but differ considerably in RNA sequence (53). Thus, riboswitches for flavin mononucleotide (FMN), fluoride, lysine, purine, thiamine pyrophosphate (TPP) and pre-queuosine 1 (preQ1) are, by prediction, encoded by the genome of *L. lactis* (119). In addition, various T-box structures have been reported that are involved in the regulation of aminoacyl-tRNA synthetases (this thesis, Chapter 2; (120)).

Many sRNAs in Gram-negative bacteria require the RNA binding protein Hfq for proper functioning. This donut-shaped heterohexamer acts as an RNA chaperone that can bind both sRNA and mRNA. Deletion of Hfq in *E. coli* leads to pleiotropic effects and increased sensitivity to some stresses (121). In *Staphylococcus aureus*, however, Hfq does not seem to be crucial for stress tolerance (122). Hfq is present in nearly 50% of the bacteria that have been sequenced, especially in bacteria with a high GC content (123). Recently, a study using gradient sequencing (Grad-Seq) and CLIP-Seq has identified ProQ as another important RNA binding protein binding and stabilizing many sRNAs (85). ProQ contains the functional FinO domain for RNA binding (124). In *L. lactis*, neither Hfq nor ProQ homologs are present. This suggests that sRNAs, specifically those that are *trans* encoded, either do not require an RNA chaperone or use another RNA binding protein(s). A protein that could potentially function as an sRNA chaperone in *L. lactis* is Llmg_1487. This protein has a predicted metalloprotease function and is homologous to *E. coli* YbeY. The YbeY homologue of the plant symbiont *Sinorhizobium meliloti* was recently identified as a potential candidate involved in sRNA regulation since a deletion mutant of this protein has a phenotype similar to that of an *E. coli*

Hfq deletion strain in that it has a reduced ability to regrow after different stress conditions (125). It has been shown that *E. coli* YbeY modulates many sRNA-mRNA interactions (126). However, other work has shown that YbeY is a ribonuclease responsible for ribosome maturation and fidelity (127-130). *L. lactis llmg_1487* is an essential gene as a knock-out strain could not be obtained while in a conditional knock-out strain carrying *llmg_1487* on a temperature-sensitive plasmid, the latter could not be cured (unpublished results). It is, thus, tempting to speculate that Llmg_1487 is involved in sRNA regulation in *L. lactis* but this hypothesis needs further investigation. RNases are reported to play an important role in the field of regulatory RNAs. In *L. lactis*, PNPase and RNase III homologs could be identified on the basis of protein sequence similarities. *E. coli* RNase E is represented in *L. lactis* by the functional paralog RNase Y. Altogether, the review presented here provides an extensive incentive to study the occurrence and function of sRNAs in *L. lactis* and related lactococcal species. It is exactly this that has been undertaken in this thesis work:

Outline

This thesis is structured in seven chapters. It starts with **Chapter 1**, which gives a general introduction on the different kinds of gene regulation, in particular those that are based on RNA, which exist in bacteria. The crucial technique in current RNA research is RNA sequencing (RNA-seq). This method is briefly presented here, including some variants of RNA-seq. The model organism *Lactococcus lactis* used in this study is described, together with the knowledge available on RNA regulation in this organism.

In **Chapter 2**, the transcriptome landscape of *L. lactis* subsp. *cremoris* MG1363 is extensively explored by analyzing the data obtained by differential RNA sequencing. A combination of intensive manual and automated transcriptome mining resulted in a high-resolution genome annotation of *L. lactis* and the identification of 60 *cis*-encoded antisense RNAs (asRNAs), 186 *trans*-encoded putative regulatory RNAs (sRNAs) and 134 new small ORFs. Fourteen sRNA and asRNA candidates were validated by Northern analysis. Functional analyses were performed on two of the sRNAs. For one of these, 6S, expression was shown to be strongly dependent of the available carbon source. Another sRNA, LLMGnc_147, was proven to activate an operon involved in the utilization and uptake of galactose.

Chapter 3 provides insights in the primary transcriptomic response of *L. lactis* NCDO712 when it was only briefly (5 min) exposed to a variety of industrially relevant stress conditions. Specific attention is paid to the expression of the 186 sRNAs that were discovered and described in Chapter 2. In addition, tRNA dynamics is monitored upon the exposure of the cells to the various stress conditions.

Chapter 4 reports in detail on the function of a small regulatory RNA called ArgX and its role

in arginine metabolism. ArgX is transcribed in the 3' untranslated region of *argR*, a gene encoding the ArgR protein regulator controlling the transcription of the arginine catabolic *arc* and anabolic *arg* operons. It was shown that ArgX, in addition to ArgR, regulates *arc*. It does so by base pairing with a 12-bp complementary region covering the RBS of *arcC1*, resulting in a downregulation of *arc*.

In **Chapter 5**, targets of the sRNA CisR , for Cold induced sRNA, were identified by overexpressing its gene and using MAPS technology, followed by RNA-seq. CisR is upregulated upon 5 min of cold shock at 10°C. A specific decrease of tRNA expression was observed in cells that underwent either a pulse-expression of CisR or a cold shock for 5 min at 10°C. Furthermore, a pull down experiment using *ms2-CisR*, a CisR variant extended at its 5'-end with an aptamer recognized by the RNA binding protein MS2, and a Maltose Binding Protein-MS2 fusion protein (MBP-MS2), resulted in the enrichment of several tRNAs relative to the control, the *ms2* aptamer alone.

Chapter 6 describes T-Rex, a user-friendly webserver tool that was used throughout this thesis work for the analyses of RNA-seq transcriptome data. This software requires only basic input from the user and runs in a matter of minutes. T-Rex performs statistical analyses and provides a wide choice of data plots, tables and figures. The output includes *e.g.* correlation matrices, k-means clusters, heatmaps and volcano plots that subsequently can be further examined by the user for biological relevance.

Finally, **Chapter 7** summarizes the general results and discusses the findings obtained in this thesis work and presents leads for future research.

REFERENCES

1. **Piggot PJ, Hilbert DW.** 2004. Sporulation of *Bacillus subtilis*. Curr Opin Microbiol **7:**579-586.
2. **Harshey RM.** 2003. Bacterial motility on a surface: many ways to a common goal. Annual Reviews in Microbiology **57:**249-273.
3. **Filloux A.** 2010. A variety of bacterial pili involved in horizontal gene transfer. J Bacteriol **192:**3243-3245.
4. **Smits WK, Kuipers OP, Veening J.** 2006. Phenotypic variation in bacteria: the role of feedback regulation. Nature Reviews Microbiology **4:**259-271.
5. **Dubnau D, Losick R.** 2006. Bistability in bacteria. Mol Microbiol **61:**564-572.
6. **de Jong IG, Haccou P, Kuipers OP.** 2011. Bet hedging or not? A guide to proper classification of microbial survival strategies. Bioessays **33:**215-223.
7. **Veening JW, Stewart EJ, Berngruber TW, Taddei F, Kuipers OP, Hamoen LW.** 2008. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc Natl Acad Sci U S A **105:**4393-4398.
8. **Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R.** 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: σS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol **187:**1591-1603.
9. **Ades SE.** 2004. Control of the alternative sigma factor σE in *Escherichia coli*. Curr Opin Microbiol **7:**157-162.
10. **Arsène F, Tomoyasu T, Bukau B.** 2000. The heat shock response of *Escherichia coli*. Int J Food Microbiol **55:**3-9.
11. **Wade JT, Roa DC, Grainger DC, Hurd D, Busby SJ, Struhl K, Nudler E.** 2006. Extensive functional overlap between σ factors in *Escherichia coli*. Nature structural & molecular biology **13:**806-814.
12. **Boyes J, Bird A.** 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell **64:**1123-1134.
13. **Waters LS, Storz G.** 2009. Regulatory RNAs in bacteria. Cell **136:**615-628.
14. **Narberhaus F, Vogel J.** 2009. Regulatory RNAs in prokaryotes: here, there and everywhere. Mol Microbiol **74:**261-269.
15. **Livny J, Waldor MK.** 2007. Identification of small RNAs in diverse bacterial species. Curr Opin Microbiol **10:**96-101.
16. **Mizuno T, Chou M, Inouye M.** 1983. Regulation of gene expression by a small RNA transcript (micRNA) in *Escherichia coli* K-12. Proc Japan Acad , Ser B **59:**335-338.
17. **Mizuno T, Chou MY, Inouye M.** 1984. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci U S A **81:**1966-1970.
18. **Cech TR, Steitz JA.** 2014. The noncoding RNA revolution—trashing old rules to forge new ones. Cell **157:**77-94.
19. **Morita T, Aiba H.** 2011. RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. Genes Dev **25:**294-298.
20. **Bandyra KJ, Said N, Pfeiffer V, Górna MW, Vogel J, Luisi BF.** 2012. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol Cell **47:**943-953.

21. **Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J.** 2013. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. Cell **153:**426-437.
22. **Busch A, Richter AS, Backofen R.** 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics **24:**2849-2856.
23. **Kery MB, Feldman M, Livny J, Tjaden B.** 2014. TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res **42:**W124-W129.
24. **Eggenhofer F, Tafer H, Stadler PF, Hofacker IL.** 2011. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res **39:**W149-54.
25. **Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J.** 2013. Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci U S A **110:**E3487-96.
26. **Papenfort K, Podkaminski D, Hinton JC, Vogel J.** 2012. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. Proc Natl Acad Sci U S A **109:**E757-64.
27. **Hoe C, Raabe CA, Rozhdestvensky TS, Tang T.** 2013. Bacterial sRNAs: regulation in stress. International Journal of Medical Microbiology **303:**217-229.
28. **Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, Thompson KM, FitzGerald PC, Sowa NA, FitzGerald DJ.** 2006. Small RNA regulators and the bacterial response to stress. Cold Spring Harb Symp Quant Biol **71:**1-11.
29. **Holmqvist E, Wagner EGH.** 2017. Impact of bacterial sRNAs in stress responses. Biochem Soc Trans **45:**1203-1212.
30. **Lioliou E, Romilly C, Geissmann T, Vandenesch F, Romby P.** 2012. RNA-mediated Regulation of Virulence Gene Expression: Another Layer of Complexity. Bacterial Regulatory Networks 143.
31. **Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, Johansson J.** 2010. RNAs: regulators of bacterial virulence. Nature Reviews Microbiology **8:**857-866.
32. **Sonnleitner E, Haas D.** 2011. Small RNAs as regulators of primary and secondary metabolism in *Pseudomonas* species. Appl Microbiol Biotechnol **91:**63-79.
33. **Storz G, Vogel J, Wassarman KM.** 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell **43:**880-891.
34. **Masse E, Gottesman S.** 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci U S A **99:**4620-4625.
35. **Bobrovskyy M, Vanderpool CK.** 2014. The small RNA SgrS: Roles in metabolism and pathogenesis of enteric bacteria. Frontiers in Cellular and Infection Microbiology **4:**61.
36. **Wadler CS, Vanderpool CK.** 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci U S A **104:**20454-20459.
37. **Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, Vogel J.** 2006. σE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. Mol Microbiol **62:**1674-1688.
38. **Georg J, Hess WR.** 2011. cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev **75:**286-300.
39. **Tomizawa J, Itoh T, Selzer G, Som T.** 1981. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. Proc Natl Acad Sci U S A **78:**1421-1425.

40. **Stougaard P, Molin S, Nordstrom K.** 1981. RNAs involved in copy-number control and incompatibility of plasmid R1. Proc Natl Acad Sci U S A **78:**6008-6012.
41. **Wagner EGH, Altuvia S, Romby P.** 2002. 12 Antisense RNAs in bacteria and their genetic elements. Adv Genet **46:**361-398.
42. **Thomason MK, Storz G.** 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet **44:**167-188.
43. **Mars RA, Mendonça K, Denham EL, van Dijl JM.** 2015. The reduction in small ribosomal subunit abundance in ethanol-stressed cells of *Bacillus subtilis* is mediated by a SigB-dependent antisense RNA. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1853:**2553-2559.
44. **Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de los Mozos IR, Vergara-Irigaray M, Segura V, Fagegaltier D, Penades JR, Valle J, Solano C, Gingeras TR.** 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. Proc Natl Acad Sci U S A **108:**20172-20177.
45. **Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G.** 2015. Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. J Bacteriol **197:**18-28.
46. **Hindley Jt.** 1967. Fractionation of 32 P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. J Mol Biol **30:**125-136.
47. **Wassarman KM, Storz G.** 2000. 6S RNA Regulates *E. coli* RNA Polymerase Activity. Cell **101:**613-623.
48. **Barrick JE, Sudarsan N, Weinberg Z, Ruzzo WL, Breaker RR.** 2005. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA **11:**774-784.
49. **Cavanagh AT, Wassarman KM.** 2014. 6S RNA, a Global Regulator of Transcription in *Escherichia coli*, Bacillus subtilis, and Beyond. Annu Rev Microbiol **68:**45-60.
50. **Babitzke P, Romeo T.** 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol **10:**156-163.
51. **Timmermans J, Van Melderen L.** 2010. Post-transcriptional global regulation by CsrA in bacteria. Cellular and molecular life sciences **67:**2897-2908.
52. **Göpel Y, Khan MA, Görke B.** 2014. Ménage à trois: Post-transcriptional control of the key enzyme for cell envelope synthesis by a base-pairing small RNA, an RNase adaptor protein, and a small RNA mimic. RNA biology **11:**433-442.
53. **Breaker RR.** 2012. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol **4:**10.1101/cshperspect.a003566.
54. **Serganov A, Nudler E.** 2013. A decade of riboswitches. Cell **152:**17-24.
55. **Green NJ, Grundy FJ, Henkin TM.** 2010. The T box mechanism: tRNA as a regulatory molecule. FEBS Lett **584:**318-324.
56. **Kortmann J, Narberhaus F.** 2012. Bacterial RNA thermometers: molecular zippers and switches. Nature Reviews Microbiology **10:**255-265.
57. **Narberhaus F, Waldminghaus T, Chowdhury S.** 2006. RNA thermometers. FEMS Microbiol Rev **30:**3-16.
58. **De Lay N, Gottesman S.** 2012. RNase E Finds Some sRNAs Stimulating. Mol Cell **47:**825-826.
59. **Viegas SC, Arraiano CM.** 2008. Regulating the regulators: how ribonucleases dictate the rules in the control of small non-coding RNAs. RNA biology **5:**230-243.
60. **Jester BC, Romby P, Lioliou E.** 2012. When ribonucleases come into play in pathogens: a survey

of gram-positive bacteria. International journal of microbiology **2012:**.

61. **Saramago M, Bárria C, dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC, Arraiano CM.** 2014. The role of RNases in the regulation of small RNAs. Curr Opin Microbiol **18:**105-115.
62. **Saramago M, Bárria C, dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC, Arraiano CM.** 2014. The role of RNases in the regulation of small RNAs. Curr Opin Microbiol **18:**105-115.
63. **Andrade JM, Arraiano CM.** 2008. PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. RNA **14:**543-551.
64. **De Lay N, Gottesman S.** 2011. Role of polynucleotide phosphorylase in sRNA function in Escherichia coli. RNA **17:**1172-1189.
65. **Ha M, Kim VN.** 2014. Regulation of microRNA biogenesis. Nature reviews Molecular cell biology **15:**509-524.
66. **Huntzinger E, Izaurralde E.** 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nature Reviews Genetics **12:**99-110.
67. **Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK.** 2003. RNA interference: biology, mechanism, and applications. Microbiol Mol Biol Rev **67:**657-685.
68. **Mercer TR, Dinger ME, Mattick JS.** 2009. Long non-coding RNAs: insights into functions. Nature Reviews Genetics **10:**155-159.
69. **Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M,**

Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. 2001. The sequence of the human genome. Science **291:**1304-1351.

70. **Wang Z, Gerstein M, Snyder M.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics **10:**57-63.
71. **Sharma CM, Vogel J.** 2014. Differential RNA-seq: the approach behind and the biological insight gained. Curr Opin Microbiol **19:**97-105.
72. **Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiß S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R.** 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464:**250-255.
73. **Ettwiller L, Buswell J, Yigit E, Schildkraut I.** 2015. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. bioRxiv 034785.
74. **Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J.** 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J **31:**4005-4019.
75. **Darnell RB.** 2010. HITS-CLIP: panoramic views of protein–RNA regulation in living cells. Wiley Interdisciplinary Reviews: RNA **1:**266-286.
76. **Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS.** 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science **324:**218-223.
77. **Shendure J, Ji H.** 2008. Next-generation DNA sequencing. Nat Biotechnol **26:**1135-1145.
78. **Metzker ML.** 2010. Sequencing technologies—the next generation. Nature reviews genetics **11:**31-46.
79. **Laird PW.** 2010. Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics **11:**191-203.
80. **Johnson DS, Mortazavi A, Myers RM, Wold B.** 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science **316:**1497-1502.
81. **van Opijnen T, Bodi KL, Camilli A.** 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature methods **6:**767-772.
82. **Innocenti N, Golumbeanu M, Fouquier d'Herouel A, Lacoux C, Bonnin RA, Kennedy SP, Wessner F, Serror P, Bouloc P, Repoila F, Aurell E.** 2015. Whole-genome mapping of 5′ RNA ends in bacteria by tagged sequencing: a comprehensive view in *Enterococcus faecalis*. RNA **21:**1018-1030.
83. **Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E.** 2010. Genome-wide measurement of RNA secondary structure in yeast. Nature **467:**103-107.
84. **Mamanova L, Andrews RM, James KD, Sheridan EM, Ellis PD, Langford CF, Ost TW, Collins JE, Turner DJ.** 2010. FRT-seq: amplification-free, strand-specific transcriptome sequencing. Nature methods **7:**130-132.
85. **Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J.** 2016. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proceedings of the National Academy of Sciences 201609981.

86. **Westermann AJ, Gorski SA, Vogel J.** 2012. Dual RNA-seq of pathogen and host. Nature Reviews Microbiology **10:**618-630.
87. **Le TB, Imakaev MV, Mirny LA, Laub MT.** 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. Science **342:**731-734.
88. **Han K, Tjaden B, Lory S.** 2016. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. Nature microbiology **2:**16239.
89. **Lalaouna D, Massé E.** 2015. Identification of sRNA interacting with a transcript of interest using MS2-affinity purification coupled with RNA sequencing (MAPS) technology. Genomics data **5:**136-138.
90. **Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS.** 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature **505:**701-705.
91. **Kang Y, Norris MH, Zarzycki-Siek J, Nierman WC, Donachie SP, Hoang TT.** 2011. Transcript amplification from single bacterium for transcriptome analysis. Genome Res **21:**925-935.
92. **Cahová H, Winz M, Höfer K, Nübel G, Jäschke A.** 2014. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature **519:**374–377.
93. **Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R.** 2014. The double-stranded transcriptome of *Escherichia coli*. Proc Natl Acad Sci U S A **111:**3134-3139.
94. **McAuliffe O.** 2018. Symposium review: *Lactococcus lactis* from nondairy sources: Their genetic and metabolic diversity and potential applications in cheese. J Dairy Sci .
95. **Gasson MJ.** 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. J Bacteriol **154:**1-9.
96. **Kuipers OP, de Ruyter PG, Kleerebezem M, de Vos WM.** 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol **64:**15-21.
97. **Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, Evershed RP.** 2013. Earliest evidence for cheese making in the sixth millennium BC in northern Europe. Nature **493:**522-525.
98. **Smit G, Smit BA, Engels WJ.** 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. FEMS Microbiol Rev **29:**591-610.
99. **Rijnen L, Yvon M, van Kranenburg R, Courtin P, Verheul A, Chambellon E, Smit G.** 2003. Lactococcal aminotransferases AraT and BcaT are key enzymes for the formation of aroma compounds from amino acids in cheese. Int Dairy J **13:**805-812.
100. **Liu M, Nauta A, Francke C, Siezen RJ.** 2008. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. Appl Environ Microbiol **74:**4590-4600.
101. **Marilley L, Casey M.** 2004. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. Int J Food Microbiol **90:**139-159.
102. **Kunji ER, Slotboom D, Poolman B.** 2003. *Lactococcus lactis* as host for overproduction of functional membrane proteins. Biochimica et Biophysica Acta (BBA)-Biomembranes **1610:**97-108.
103. **Nouaille S, Ribeiro LA, Miyoshi A, Pontes D, Le Loir Y, Oliveira SC, Langella P, Azevedo V.** 2003. Heterologous protein production and delivery systems for *Lactococcus lactis*. Genet Mol Res **2:**102-111.
104. **Kuipers OP, Bierbaum G, Ottenwälder B, Dodd HM, Horn N, Metzger J, Kupke T, Gnau V, Bongers R, van den Bogaard P.** 1996. Protein engineering of lantibiotics. Antonie Van Leeuwenhoek

69:161-170.

105. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. lactis IL1403. Genome Res **11:**731-753.

106. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J. 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. J Bacteriol **189:**3256-3270.

107. Ventura M, Zomer A, Canchaya C, O'Connell-Motherway M, Kuipers O, Turroni F, Ribbera A, Foroni E, Buist G, Wegmann U, Shearman C, Gasson MJ, Fitzgerald GF, Kok J, van Sinderen D. 2007. Comparative analyses of prophage-like elements present in two *Lactococcus lactis* strains. Appl Environ Microbiol **73:**7771-7780.

108. Kok J, van Gijtenbeek LA, de Jong A, van der Meulen, Sjoerd B, Solopova A, Kuipers OP. 2017. The Evolution of gene regulation research in *Lactococcus lactis*. FEMS Microbiol Rev **41:**S220-S243.

109. Kuipers OP, Beerthuyzen MM, de Ruyter PG, Luesink EJ, de Vos WM. 1995. Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction. J Biol Chem **270:**27299-27304.

110. Mierau I, Kleerebezem M. 2005. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. Appl Microbiol Biotechnol **68:**705-717.

111. van Pijkeren JP, Britton RA. 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Res **40:**e76.

112. Pinto JP, Zeyniyev A, Karsens H, Trip H, Lolkema JS, Kuipers OP, Kok J. 2011. pSEUDO, a genetic integration standard for *Lactococcus lactis*. Appl Environ Microbiol **77:**6687-6690.

113. Roces C, Perez V, Campelo AB, Blanco D, Kok J, Kuipers OP, Rodriguez A, Martinez B. 2012. The putative lactococcal extracytoplasmic function anti-sigma factor llmg2447 determines resistance to the cell wall-active bacteriocin lcn972. Antimicrob Agents Chemother **56:**5520-5527.

114. Zomer AL, Buist G, Larsen R, Kok J, Kuipers OP. 2007. Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. *cremoris* MG1363. J Bacteriol **189:**1366-1381.

115. Guédon E, Serror P, Ehrlich SD, Renault P, Delorme C. 2001. Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. Mol Microbiol **40:**1227-1239.

116. den Hengst CD, van Hijum SA, Geurts JM, Nauta A, Kok J, Kuipers OP. 2005. The *Lactococcus lactis* CodY regulon: identification of a conserved cis-regulatory element. J Biol Chem **280:**34332-34342.

117. Larsen R, van Hijum SA, Martinussen J, Kuipers OP, Kok J. 2008. Transcriptome analysis of the *Lactococcus lactis* ArgR and AhrC regulons. Appl Environ Microbiol **74:**4768-4771.

118. Martínez B, Zomer AL, Rodríguez A, Kok J, Kuipers OP. 2007. Cell envelope stress induced by the bacteriocin Lcn972 is sensed by the lactococcal two-component system CesSR. Mol Microbiol **64:**473-486.

119. van der Meulen SB, de Jong A, Kok J. 2016. Transcriptome landscape of *Lactococcus lactis* reveals many novel RNAs including a small regulatory RNA involved in carbon uptake and metabolism. RNA Biol **13:**353-366.

120. Wels M, Groot Kormelink T, Kleerebezem M, Siezen RJ, Francke C. 2008. An in silico analysis of T-box regulated genes and T-box evolution in prokaryotes, with emphasis on prediction of

substrate specificity of transporters. BMC Genomics **9:**330-2164-9-330.

121. Tsui HT, Leung HE, Winkler ME. 1994. Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. Mol Microbiol **13:**35-49.

122. Bohn C, Rigoulay C, Bouloc P. 2007. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. BMC microbiology **7:**10.

123. Jousselin A, Metzinger L, Felden B. 2009. On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol **17:**399-405.

124. G. Chaulk S, Smith– Frieday MN, Arthur DC, Culham DE, Edwards RA, Soo P, Frost LS, Keates RA, Glover JM, Wood JM. 2011. ProQ is an RNA chaperone that controls ProP levels in *Escherichia coli*. Biochemistry (N Y) **50:**3095-3106.

125. Pandey SP, Minesinger BK, Kumar J, Walker GC. 2011. A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. Nucleic Acids Res **39:**4691-4708.

126. Pandey SP, Winkler JA, Li H, Camacho DM, Collins JJ, Walker GC. 2014. Central role for RNase YbeY in Hfq-dependent and Hfq-independent small-RNA regulation in bacteria. BMC Genomics **15:**121.

127. Jacob AI, Köhrer C, Davies BW, RajBhandary UL, Walker GC. 2013. Conserved bacterial RNase YbeY plays key roles in 70S ribosome quality control and 16S rRNA maturation. Mol Cell **49:**427-438.

128. Davies BW, Köhrer C, Jacob AI, Simmons LA, Zhu J, Aleman LM, RajBhandary UL, Walker GC. 2010. Role of *Escherichia coli* YbeY, a highly conserved protein, in rRNA processing. Mol Microbiol **78:**506-518.

129. Rasouly A, Davidovich C, Ron EZ. 2010. The heat shock protein YbeY is required for optimal activity of the 30S ribosomal subunit. J Bacteriol **192:**4592-4596.

130. Rasouly A, Schonbrun M, Shenhar Y, Ron EZ. 2009. YbeY, a heat shock protein involved in translation in *Escherichia coli*. J Bacteriol **191:**2649-2655.